5К

УДК 81'33

О.А. Каплина, А.В. Типаков
Институт проблем искусственного интеллекта МОН и НАН Украины, г. Донецк
kaplina@iai.donetsk.ua

Принципы построения электронного «Инверсионного грамматического словаря украинского языка»

В данной статье рассматриваются основные принципы построения разработанного электронного «Инверсионного грамматического словаря украинского языка». Данный словарь предназначен для предоставления пользователю морфологической информации о слове, а также полной его парадигмы. В роли указателей основной морфологической характеристики каждого слова выступают буквенные символы – метки. Каждая метка определяет алгоритм словоизменения.

Введение

Создание украинского грамматического словаря и программ украинского автоматического словоизменения – актуальные вопросы современной компьютерной лингвистики. Кроме того, украинский грамматический словарь должен стать основой автоматического семантико-синтаксического анализатора текста. Компьютерный «Інверсійний граматичний словник української мови» предназначен для предоставления пользователю морфологической информации о слове, а также его словоизменительной характеристики, то есть его полной парадигмы.

Целью работы является создание электронного «Инверсионного грамматического словаря украинского языка». Для достижения данной цели **поставлены** и **решены** следующие **задачи**:

1. Выявление и формализация закономерностей украинского словоизменения.
2. Синтез системы меток и алгоритмов словоизменения.
3. Разметка словарной базы и тестирование алгоритмов.

Принципы построения словаря

Слова данного словаря объединены в группы по принципу принадлежности слов к определенной части речи и по общим конечным формально выраженным элементам. Каждое слово представлено в словаре начальной формой. Проставленные рядом метки содержат морфологическую информацию о данном слове. По сути, совокупность меток, принадлежащих каждой словарной единице, определяет алгоритм изменения слова.

Метка представлена буквенными символами, цифровыми индексами, а также дополнительными символами при цифре. Буквенные символы указывают основную морфологическую характеристику имени (часть речи, род, одушевленность – неодушевленность).

В роли указателей основной морфологической характеристики имени буквенные символы (основные или дополнительные) имеют следующие значения:

ч – склонение существительных мужского рода неодушевленных;

чо – склонение существительных мужского рода одушевленных;

ж – склонение существительных женского рода неодушевленных;
жо – склонение существительных женского рода одушевленных;
с – склонение существительных среднего рода неодушевленных;
со – склонение существительных среднего рода одушевленных;
пр – склонение прилагательных;
з – склонение местоимений;
ч – склонение числительных.

В словаре каждому существительному соответствует метка, которая имеет следующую структуру:

одушевленность/неодушевленность_____тип склонения___[дополнительная метка]___схема ударения.

Например:

аболіціоніст чо1а Павлівна жо1!а

Склонение существительных характеризуется определенными изменяемыми окончаниями, которые добавляются к неизменяемой основе.

Цифра индекса (от 0 до 10) обозначает группу склонения.

Цифра 0 означает, что слово неизменяемо (т.е. все формы внешне одинаковы).

Цифры 1 – 10 означают разные группы склонения, зависящие от окончания и от того, на что оканчивается основа слова. Склонение слов разных групп различаются окончаниями. Окончания основных групп приведены в таблицах. «□» означает нулевое окончание.

Звездочка при цифре индекса показывает, что при склонении в основе данного слова происходит чередование так называемой беглой гласной с нулем звука.

Всякое слово, имеющее индекс со звездочкой, отличается от слова, имеющего соответствующий индекс без звездочки, только тем, что имеет в основе чередование беглой гласной с нулем; в остальных отношениях (окончания, ударение и т.д.) такие два слова одинаковы.

Латинская буква от a до f (или две буквы, разделенные косой чертой), без штрихов или со штрихами – обозначение схемы ударения. Общее значение букв: a – постоянное ударение на основе; b – ударение на окончаниях (кроме случаев, когда окончание не содержит гласной); от c до f – разные виды подвижного ударения.

Каждому прилагательному в словаре соответствует метка, которая имеет следующую структуру:

тип склонения___степень сравнения___схема ударения.

Например:

молод>ий пр1Ав/ навігац>ійний пр1Са/

Прилагательные были распределены на пять групп в зависимости от окончаний и типа склонения. Метка А и В указывает на возможность образования сравнительной и превосходной степеней сравнения. Метка С – на отсутствие такой возможности. Способы формирования степеней сравнения для удобства представлены таблицами. Также следует обратить внимание на случаи чередования основы прилагательного при создании высшей простой степени сравнения.

В словаре каждому местоимению соответствует метка, которая обозначает один из семи типов склонения.

Например:

будь-котрий з1 будь-чий з5

Тип 1 включает в себя местоимения на **-ий**; **-хто** (в множественном числе – основа < -к->), **-що** (во множественном числе основа – основа <-ч->); **-той** (параллельные формы женского рода только для местоимения на -той); **наш**, **ваш** и в косвенных падежах имеет окончания прилагательного первого типа. Тип 2 характеризуется окончаниями **-ій** (кроме «їхній»), **мій**, **твій**, **свій**. Тип 3 имеет окончание – **ей** и также характеризуется определенной системой окончаний. К типу 4 относится слово «їхній», которое склоняется по образцу прилагательного типа 2. Тип 5 включает в себя слова на -чий и имеет общие окончания в косвенных падежах. Тип 6 включает в себя такие слова, как «я» (в косвенных падежах единственного числа – основа «мен», множественного – «н»); «ти» (в косвенных падежах: единственного числа – основа «мен», множественного – «в»); «себе» (нет формы именительного падежа; основа – «себ»; нет множественного числа). Тип 7 состоит из слова «він», форм женского и среднего рода данного местоимения и формы его множественного числа.

В словаре каждому глаголу соответствует метка, которая имеет следующую структуру:

вид глагола___тип спряжения___схема ударения___ [дополнительная метка] ___ переходность ____приставка.

Например:

без>умствувати ндв 2а нп +   буркнути дв 3а

Вид глагола обозначается русскими буквами: дв (совершенный вид) или ндв (несовершенный вид), и ндв-дв в случае двувидового глагола;

Тип спряжения обозначается цифрой от 1 до 18 и буквосочетанием CAPut!’. Каждый тип спряжения характеризуется определенным способом построения 3 основных форм, от которых строятся все остальные формы парадигмы. Это: 1) инфинитив; 2) 1 лицо единственного числа настоящего времени (для глаголов несовершенного вида) или будущего (для глаголов совершенного вида); 3) 3 лицо единственного числа настоящего (будущего) времени. Далее будем условно называть – I, II или III форма глагола.

Схема ударения обозначается латинской буквой (или двумя буквами, разделенными косой чертой). Схема ударения складывается из 2 частей: 1) схема ударения личных форм настоящего (будущего) времени и повелительного наклонения (сокращенно – схема ударения наст. времени, ее символ пишется слева от косой черты); 2) схема ударения личных форм прошедшего времени (ее символ пишется справа от косой черты). При этом метка *a*, выступающая как 2-й элемент записи, опускается, т.е. вместо *a/a* пишется *a*, вместо *b/a* – *b*, вместо *c/a* – *c*.

«Переходность» – переходность глагола, если глагол является непереходным, то ставится метка нп; отсутствие такой метки говорит о том, что глагол является переходным.

«Приставка» – метка «+» означает наличие приставки в данном глаголе, она введена из-за сложности автоматического определения наличия приставки, что необходимо при построении причастий.

«Дополнительная метка» – дополнительные метки, обозначающие различные особенности при образовании парадигм (напр., чередование в приставке). Квадратные скобки означают, что этих меток может и не быть.

«*» – показывает, что при спряжении происходит чередование беглой гласной с нулем в конце приставки (від-/віді-, в-/вві-, над-/наді-, роз-/розі- и др.).

«@» – означает, что слово имеет особенности (неправильности) при спряжении (брати – беру).

«#» – для глаголов с ударной приставкой «ви-», которая влияет на схему ударения, но повелительное наклонение образуется так же, как и в глаголах без этой приставки со схемой “b” или “c”.

«^» – означает: для типа 6 отсутствие стандартного чередования согласных.

Типы 4 и 5 соответствуют I спряжению глагола, все остальные – II спряжению. Только в архаичном типе представлены оба спряжения, словоформы этого типа образуются особым образом.

Особого внимания требует построение форм повелительного наклонения. Так, форма 2 лица единственного числа повелительного наклонения имеет ту же основу, что и в 3 лице настоящего времени (будущего) (т.е. в I спряжении основа 1 лица и во II спряжении основа 3 лица единственного числа). Исключение: в типе 13 формы повелительного наклонения образуются от инфинитива (давати – давай).

Окончания -и, -імо(-ім), -іть(-іте) соответствуют тем глаголам, которые имеют схему ударения /b или /c, когда основа заканчивается двумя согласными или стоит метка «#». В остальных случаях – чистая основа для 2 лица единственного числа и соответственно -мо и -те для других. Эти глаголы делятся на 3 группы: 1) оканчиваются на -й (читай, лай); 2) грамматическая основа заканчивается на д, т, з, с, л, н, в повелит., они смягчаются (кинуть – кинь, киньмо, киньте); 3) глаголы на согласные, которые не смягчаются.

Заключение

Созданный «Інверсійний граматичний словник української мови», который является, по сути, электронным аналогом печатного «Грамматического словаря русского языка» А.А. Зализняка, содержит около ста тысяч словарных статей. Планируется издание печатного варианта «Інверсійного граматичного словника української мови». Книга будет содержать словник, каждое слово в котором снабжено меткой, определяющей алгоритм словоизменения, также в ней будет содержаться часть, в которой объясняется, как, используя метку, реализовать алгоритм словоизменения.

Грамматический словарь и библиотеки автоматического словоизменения украинского языка могут стать основой для будущего семантико-синтаксического анализатора украиноязычных текстов. Данный словарь будет полезным как для неподготовленного говорящего, который изучает украинский литературный язык с целью общения, так и для специалистов, которые исследуют языковые явления и закономерности.

Литература

1. Загнітко А.П. Теоретична граматика української мови. Морфологія. – Донецьк: ДонДУ, 1996. – 300 с.
2. Зализняк А.А. Грамматический словарь русского языка: Словоизменение. – М.: Рус. яз., 1977. – 879 с.
3. Зализняк А.А. Русское именное словоизменение. – М.: Наука, 1967. – 310 с.
4. Филлмор Ч. Дело о падеже // Новое в зарубежной лингвистике. – Вып. 10. – М.: Прогресс, 1981. – С. 369-495.

О.О. Капліна, А.В. Тіпаков

Принципи побудови електронного «Інверсійного граматичного словника української мови»

У даній статті розглядаються основні принципи побудови розробленого електронного «Інверсійного граматичного словника української мови». Даний словник призначено для надання користувачу морфологічної інформації про слово, а також повної його парадигми. В ролі вказівників основної морфологічної характеристики кожного слова виступають літерні символи – мітки. Кожна мітка визначає алгоритм словозміни.

O.A. Kaplina, A.V. Tipakov

Principles of Construction of Electronic «The Inversion Grammatical Dictionary of the Ukrainian Language»

The main principles of construction electronic «The inversion grammatical dictionary of the Ukrainian language» are considered in the given article. The dictionary is to provide for user morphological information about a word, and also its full paradigm. Alphabetic symbols – labels act as indexes of the basic morphological characteristics of each word. Each label defines algorithm of word change.

Статья поступила в редакцию 16.07.2008.